Chapter 2 - Probability

2.1 - Introduction

• In everyday conversation, the term probability is a measure of one's belief in the occurrence of a future event

→ Random events, or stochastic events [cannot be predicted with certainty]
eg: Load of a bridge before collapse.

2.2 - Probability & inference

* improbable ≠
* impossible

2.3 - A Review of Set Notation

- $A, B, C, \ldots \rightsquigarrow$ set of points
- $a_1, a_2, a_3 \ldots \rightsquigarrow$ elements in the set A.

$\therefore A = \{a_1, a_2, a_3\}$.

- $S \rightsquigarrow$ set of all elements

↳ $A \subset B \rightsquigarrow A$ is a subset of B

(if every point of A is also in B.

- $\phi \rightsquigarrow$ null/empty set.

[set consisting of no points]

$\therefore \phi$ is a subset of every set.

Note: $A \cup B$ (A or B)

$A \cap B$ (A and B)

- Disjoint or mutually exclusive [$A \cap B = \phi$]

Figure 2.6

Examples of Venn Diagrams

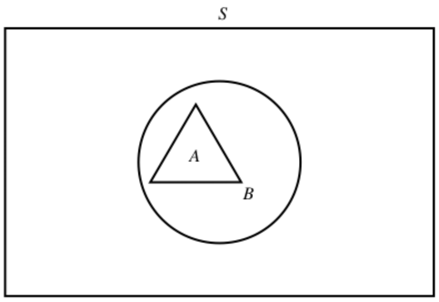


FIGURE **2.2**
Venn diagram for $A \subset B$

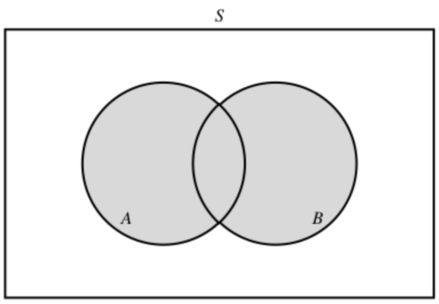


FIGURE **2.3**
Venn diagram for $A \cup B$

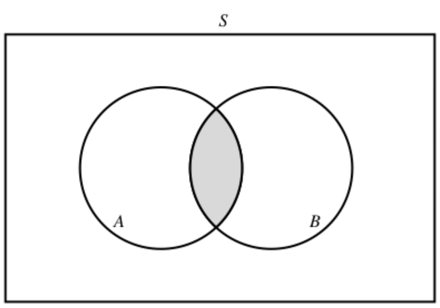


FIGURE **2.4**
Venn diagram for AB

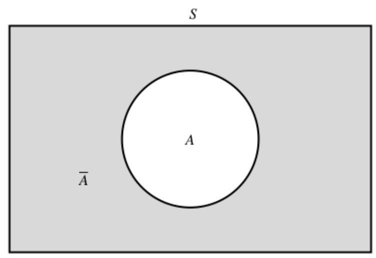


FIGURE **2.5**
Venn diagram for $\overline{A}$

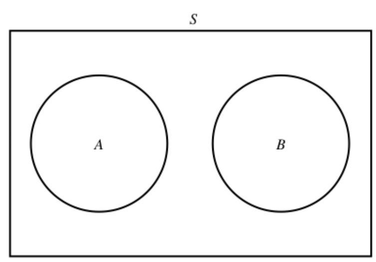


FIGURE **2.6**
Venn diagram for mutually exclusive sets A and B

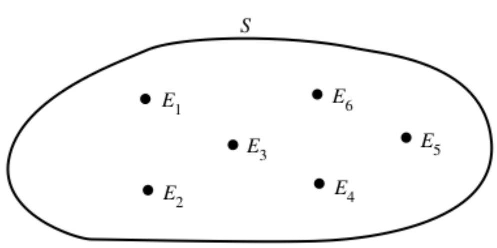
2.4 ~ A Probabilistic Model for an Experiment
The Discrete Case

* experiment: observations obtained from completely uncontrollable situations; [such as observations on the daily price of a particular stock]

* A simple event is an event that cannot be decomposed. Each simple event corresponds to one and only one sample point.

eg) letter E_n; $n = 1, 2, 3, 4 \ldots$; will be used to denote a simple event or the corresponding sample point

FIGURE 2.7
Venn diagram for the sample space associated with the die-tossing experiment

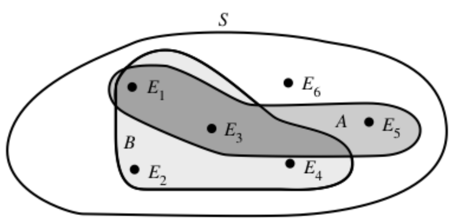


* Sample space: set consisting of all the possible sample points. [denoted by S]

→ A discrete sample space is one that contains either a finite or a countable number of distinct sample points.

* Note: All distinct simple events correspond to mutually exclusive sets of simple events, and are thus mutually exclusive events

FIGURE **2.8**
Venn diagram for the die-tossing experiment



events E_1, E_3, or E_5 occurs. Thus,

$$A = \{E_1, E_3, E_5\} \quad \text{or} \quad A = E_1 \cup E_3 \cup E_5.$$

Similarly, B (observe a number less than 5) can be written as

$$B = \{E_1, E_2, E_3, E_4\} \quad \text{or} \quad B = E_1 \cup E_2 \cup E_3 \cup E_4.$$

The rule for determining which simple events to include in a compound event is very precise. *A simple event E_i is included in event A if and only if A occurs whenever E_i occurs.*

- An event in a discrete sample space S is a collection of sample points.
  ↳ subset of S

- Analysing the relative frequency concept
  ↳ {3 AXIOMS}

DEFINITION **2.6**

Suppose S is a sample space associated with an experiment. To every event A in S (A is a subset of S), we assign a number, $P(A)$, called the *probability* of A, so that the following axioms hold:

Axiom 1: $P(A) \geq 0$.
Axiom 2: $P(S) = 1$.
Axiom 3: If $A_1, A_2, A_3, \ldots$ form a sequence of pairwise mutually exclusive events in S (that is, $A_i \cap A_j = \emptyset$ if $i \neq j$), then

$$P(A_1 \cup A_2 \cup A_3 \cup \cdots) = \sum_{i=1}^{\infty} P(A_i).$$

Note : AXIOM 3
If 2 events are mutually exclusive the relative frequency of their union is the sum of their respective relative frequencies

eg If tossing a balance die yields a 1 is = 1/6, It should yield 1 or 2
→ 1/6 + 1/6 = 1/3 of the tosses

Note:

The definition states only the conditions an assignment of probabilities must satisfy; it does not tell us how to assign specific probabilities to events.

Example: #

Suppose that a coin has yielded 800 heads in 1000 previous tosses.

Consider the experiment for one more toss of the coin.

↳ 2 possible outcomes (Heads or Tails)

Hence 2 simple events

> The definition of probability allows us to assign to these simple events any 2 non-negative numbers that add to 1

> For example each simple event could have the probability 1/2

> Taking through the history of that coin, it will be reasonable to assign a probability nearer to 0.8 to the outcome involving a head.

∴ Specific assignments of probabilities must be consistent with reality, if the probabilistic model is to serve a useful purpose

Note ②:

It is also reasonable to assume that all simple events would have the same relative frequency in the long run

∴ $P(E_i) = 1/6$ for $i = 1, 2, \ldots, 6$.

↳ Agrees of AXIOM ①

22

$$P(S) = P(E_1 \cup E_2 \cup E_3 \ldots \cup E_6) =$$

$$= P(E_1) + P(E_2) + P(E_3) \ldots + P(E_6)$$

$$= 1$$

∴ AXIOM 2 follows because AXIOM 3 holds.

Note 3: AXIOM 3 also tells us that we can calculate the probability of any event by summing the probabilities of the simple events in that event

2.5 - Calculating the Probability of an Event: Sampling Point Method

2 methods
↳ 1) Sample point
↳ 2) Event-composition methods

Both methods use the sample space model but, they differ in the sequence of steps

The sample-point method is outlined in Section 2.4. The following steps are used to find the probability of an event:

1. Define the experiment and clearly determine how to describe one simple event.
2. List the simple events associated with the experiment and test each to make certain that it cannot be decomposed. This defines the sample space S.
3. Assign reasonable probabilities to the sample points in S, making certain that $P(E_i) \geq 0$ and $\sum P(E_i) = 1$.
4. Define the event of interest, A, as a specific collection of sample points. (A sample point is in A if A occurs when the sample point occurs. Test *all* sample points in S to identify those in A.)
5. Find $P(A)$ by summing the probabilities of the sample points in A.

Example (pg 63/939)

EXAMPLE **2.4** The odds are two to one that, when A and B play tennis, A wins. Suppose that A and B play two matches. What is the probability that A wins at least one match?

Solution

1. The experiment consists of observing the winner (A or B) for each of two matches. Let AB denote the event that player A wins the first match and player B wins the second.
2. The sample space for the experiment consists of four sample points:

$$E_1: AA, \quad E_2: AB, \quad E_3: BA, \quad E_4: BB$$

3. Because A has a better chance of winning any match, it does not seem appropriate to assign equal probabilities to these sample points. As you will see in Section 2.9, under certain conditions it is reasonable to make the following assignment of probabilities:

$$P(E_1) = 4/9, \qquad P(E_2) = 2/9, \qquad P(E_3) = 2/9, \qquad P(E_4) = 1/9.$$

Notice that, even though the probabilities assigned to the simple events are not all equal, $P(E_i) \geq 0$, for $i = 1, 2, 3, 4$, and $\sum_S P(E_i) = 1$.

4. The event of interest is that A wins at least one game. Thus, if we denote the event of interest as C, it is easily seen that

$$C = E_1 \cup E_2 \cup E_3.$$

5. Finally,

$$P(C) = P(E_1) + P(E_2) + P(E_3) = 4/9 + 2/9 + 2/9 = 8/9. \qquad \blacksquare$$

2-6 – Tools for Counting Sample Points

* Useful results of the theory of combinatorial analysis & illustrate their application, to the sample-point method for finding the probability of an event

→ When the # of simple events in a sample space is very large ① & manual ② enumeration of every sample point is tedious, or even impossible.

→ Counting the number of points in the sample space and in the event of interest, may be the only efficient way to calculate the probability of an event

> If sample space contains N equiprobable sample points & an event A contains na sample points

$\Rightarrow \mathbb{P}(A) = n_a/N$

> The first result from combinatorial analysis → often called mn rule

THEOREM 2.1 With m elements $a_1, a_2, \ldots, a_m$ and n elements $b_1, b_2, \ldots, b_n$, it is possible to form $mn = m \times n$ pairs containing one element from each group.

Proof Verification of the theorem can be seen by observing the rectangular table in Figure 2.9. There is one square in the table for each a_i, b_j pair and hence a total of $m \times n$ squares.

The mn rule can be extended to any number of sets. Given three sets of elements—$a_1, a_2, \ldots, a_m$; $b_1, b_2, \ldots, b_n$; and $c_1, c_2, \ldots, c_p$—the number of distinct triplets containing one element from each set is equal to mnp. The proof of the theorem for three sets involves two applications of Theorem 2.1. We think of the first set as an (a_i, b_j) pair and unite each of these pairs with elements of the third set, $c_1, c_2, \ldots, c_p$. Theorem 2.1 implies that there are mn pairs (a_i, b_j). Because there are p elements $c_1, c_2, \ldots, c_p$, another application of Theorem 2.1 implies that there are $(mn)(p) = mnp$ triplets $a_i b_j c_k$.

Example:

- Tossing a pair of dice → observing # on the upper faces.

→ Find the number of sample point in S

→ Find the sample space for the experiment

> Sample space S consist of the sets of all possible pairs (x, y) where x & y between 1 & 6

> The first & second die both result in one of six numbers

$\therefore$ Die (1) $= a_1, a_2, a_3, \ldots, a_6$

Die (2) $= b_1, b_2, b_3 \ldots, b_6$

$\Rightarrow$ Then $m \times n = 6 \times 6 = 36$

↳ Total number of sample point in S

* We have seen that the sample points associated with an experiment often can be represented symbolically as a sequence of numbers or symbols.

↳ In some instances, it will be clear that the total number of sample points equals the number of distinct ways that the respective symbols can be arranged in sequence

DEFINITION 2.7 An ordered arrangement of r distinct objects is called a *permutation*. The number of ways of ordering n distinct objects taken r at a time will be designated by the symbol P_r^n.

THEOREM 2.2

$$P_r^n = n(n-1)(n-2)\cdots(n-r+1) = \frac{n!}{(n-r)!}.$$

Proof We are concerned with the number of ways of filling r positions with n distinct objects. Applying the extension of the mn rule, we see that the first object can be chosen in one of n ways. After the first is chosen, the second can be chosen in $(n-1)$ ways, the third in $(n-2)$, and the rth in $(n-r+1)$ ways. Hence, the total number of distinct arrangements is

$$P_r^n = n(n-1)(n-2)\cdots(n-r+1).$$

Expressed in terms of factorials,

$$P_r^n = n(n-1)(n-2)\cdots(n-r+1)\frac{(n-r)!}{(n-r)!} = \frac{n!}{(n-r)!}$$

where $n! = n(n-1)\cdots(2)(1)$ and $0! = 1$.

EXAMPLE 2.8 The names of 3 employees are to be randomly drawn, without replacement, from a bowl containing the names of 30 employees of a small company. The person whose name is drawn first receives \$100, and the individuals whose names are drawn second and third receive \$50 and \$25, respectively. How many sample points are associated with this experiment?

Solution Because the prizes awarded are different, the number of sample points is the number of ordered arrangements of $r = 3$ out of the possible $n = 30$ names. Thus, the number of sample points in S is

$$P_3^{30} = \frac{30!}{27!} = (30)(29)(28) = 24{,}360.$$

■

- The next result from combinatorial analysis can be used to determine the number of subsets of various sizes that can be formed by partitioning a set of n distinct objects into k non-overlapping groups

THEOREM 2.3

The number of ways of partitioning n distinct objects into k distinct groups containing $n_1, n_2, \ldots, n_k$ objects, respectively, where each object appears in exactly one group and $\sum_{i=1}^{k} n_i = n$, is

$$N = \binom{n}{n_1\ n_2\ \cdots\ n_k} = \frac{n!}{n_1!\ n_2!\ \cdots\ n_k!}.$$

Proof

N is the number of distinct arrangements of n objects in a row for a case in which rearrangement of the objects within a group does not count. For example, the letters a to l are arranged in three groups, where $n_1 = 3, n_2 = 4$, and $n_3 = 5$:

$$abc|defg|hijkl$$

is one such arrangement.

The number of distinct arrangements of the n objects, assuming all objects are distinct, is $P_n^n = n!$ (from Theorem 2.2). Then P_n^n equals the number of ways of partitioning the n objects into k groups (ignoring order within groups) multiplied by the number of ways of ordering the $n_1, n_2, \ldots, n_k$ elements within each group. This application of the extended mn rule gives

$$P_n^n = (N) \cdot (n_1!\, n_2!\, n_3! \cdots n_k!),$$

where $n_i!$ is the number of distinct arrangements of the n_i objects in group i.

Solving for N, we have

$$N = \frac{n!}{n_1!\, n_2! \cdots n_k!} \equiv \binom{n}{n_1\ n_2\ \cdots\ n_k}.$$

The terms $\binom{n}{n_1\ n_2\ \cdots\ n_k}$ are often called *multinomial coefficients* because they occur in the expansion of the *multinomial term* $y_1 + y_2 + \cdots + y_k$ raised to the nth power:

$$(y_1 + y_2 + \cdots + y_k)^n = \sum \binom{n}{n_1\ n_2\ \cdots\ n_k} y_1^{n_1} y_2^{n_2} \cdots y_k^{n_k},$$

where this sum is taken over all $n_i = 0, 1, \ldots, n$ such that $n_1 + n_2 + \cdots + n_k = n$.

Example:

EXAMPLE **2.10** A labor dispute has arisen concerning the distribution of 20 laborers to four different construction jobs. The first job (considered to be very undesirable) required 6 laborers; the second, third, and fourth utilized 4, 5, and 5 laborers, respectively. The dispute arose over an alleged random distribution of the laborers to the jobs that placed all 4 members of a particular ethnic group on job 1. In considering whether the assignment represented injustice, a mediation panel desired the probability of the observed event. Determine the number of sample points in the sample space S for this experiment. That is, determine the number of ways the 20 laborers can be divided into groups of the appropriate sizes to fill all of the jobs. Find the probability of the observed event if it is assumed that the laborers are randomly assigned to jobs.

Solution The number of ways of assigning the 20 laborers to the four jobs is equal to the number of ways of partitioning the 20 into four groups of sizes $n_1 = 6$, $n_2 = 4$, $n_3 = n_4 = 5$. Then

$$N = \binom{20}{6\ 4\ 5\ 5} = \frac{20!}{6!\,4!\,5!\,5!}.$$

By a *random assignment* of laborers to the jobs, we mean that each of the N sample points has probability equal to $1/N$. If A denotes the event of interest and n_a the number of sample points in A, the sum of the probabilities of the sample points in A is $P(A) = n_a(1/N) = n_a/N$. The number of sample points in A, n_a, is the number of ways of assigning laborers to the four jobs with the 4 members of the ethnic group all going to job 1. The remaining 16 laborers need to be assigned to the remaining jobs. Because there remain two openings for job 1, this can be done in

$$n_a = \binom{16}{2\ 4\ 5\ 5} = \frac{16!}{2!\,4!\,5!\,5!}$$

ways. It follows that

$$P(A) = \frac{n_a}{N} = 0.0031.$$

Thus, if laborers are randomly assigned to jobs, the probability that the 4 members of the ethnic group all go to the undesirable job is very small. There is reason to doubt that the jobs were randomly assigned. ■

• In many situations, the sample points are identified by an array of symbols in which the arrangement of symbols is unimportant

DEFINITION 2.8 The number of *combinations* of n objects taken r at a time is the number of subsets, each of size r, that can be formed from the n objects. This number will be denoted by C_r^n or $\binom{n}{r}$.

THEOREM 2.4 The number of unordered subsets of size r chosen (without replacement) from n available objects is

$$\binom{n}{r} = C_r^n = \frac{P_r^n}{r!} = \frac{n!}{r!(n-r)!}.$$

Proof The selection of r objects from a total of n is equivalent to partitioning the n objects into $k = 2$ groups, the r selected, and the $(n - r)$ remaining. This is a special case of the general partitioning problem dealt with in Theorem 2.3. In the present case, $k = 2$, $n_1 = r$, and $n_2 = (n - r)$ and, therefore,

$$\binom{n}{r} = C_r^n = \binom{n}{r \quad n-r} = \frac{n!}{r!(n-r)!}.$$

The terms $\binom{n}{r}$ are generally referred to as *binomial coefficients* because they occur in the *binomial expansion*

$$(x+y)^n = \binom{n}{0}x^n y^0 + \binom{n}{1}x^{n-1}y^1 + \binom{n}{2}x^{n-2}y^2 + \cdots + \binom{n}{n}x^0 y^n$$

$$= \sum_{i=0}^{n} \binom{n}{i} x^{n-i} y^i.$$

2.7 - Conditional Probability & the Independence of Events

. The probability of an event will sometimes depend upon whether we know that other events have occurred.
[conditional probability]

DEFINITION 2.9 The *conditional probability of an event* A, given that an event B has occurred, is equal to

$$P(A|B) = \frac{P(A \cap B)}{P(B)},$$

provided $P(B) > 0$. [The symbol $P(A|B)$ is read "probability of A given B."]

Further confirmation of the consistency of Definition 2.9 with the relative frequency concept of probability can be obtained from the following construction. Suppose that an experiment is repeated a large number, N, of times, resulting in both A and B, $A \cap B$, n_{11} times; A and not B, $A \cap \overline{B}$, n_{21} times; B and not A, $\overline{A} \cap B$, n_{12} times; and neither A nor B, $\overline{A} \cap \overline{B}$, n_{22} times. These results are contained in Table 2.1.

Note that $n_{11} + n_{12} + n_{21} + n_{22} = N$. Then it follows that

$$P(A) \approx \frac{n_{11} + n_{21}}{N}, \qquad P(B) \approx \frac{n_{11} + n_{12}}{N}, \qquad P(A|B), \approx \frac{n_{11}}{n_{11} + n_{12}},$$

$$P(B|A) \approx \frac{n_{11}}{n_{11} + n_{21}}, \qquad \text{and } P(A \cap B) \approx \frac{n_{11}}{N},$$

where $\approx$ is read *approximately equal to*.

With these probabilities, it is easy to see that

$$P(B|A) \approx \frac{P(A \cap B)}{P(A)} \qquad \text{and} \qquad P(A|B) \approx \frac{P(A \cap B)}{P(B)}.$$

Hence, Definition 2.9 is consistent with the relative frequency concept of probability.

Table 2.1 Table for events A and B

	A	$\overline{A}$	
B	n_{11}	n_{12}	$n_{11} + n_{12}$
$\overline{B}$	n_{21}	n_{22}	$n_{21} + n_{22}$
	$n_{11} + n_{21}$	$n_{12} + n_{22}$	N

Note:

The conditional probability of an event is the probability (relative frequency of occurrence) of the event given the fact that one or more events have already occured.

* Suppose that probability of the occurence of an event A is unaffected by the non-occurrence of B.

↳ ∴ A & B are independent.

DEFINITION **2.10**

Two events A and B are said to be *independent* if any one of the following holds:

$$P(A|B) = P(A),$$

$$P(B|A) = P(B),$$

$$P(A \cap B) = P(A)P(B).$$

Otherwise, the events are said to be *dependent*.

2.8 - 2 Laws of Probability

- The following 2 laws give the probabilities of unions & intersections of events.

THEOREM 2.5

The Multiplicative Law of Probability The probability of the intersection of two events A and B is

$$P(A \cap B) = P(A)P(B|A)$$
$$= P(B)P(A|B).$$

If A and B are independent, then

$$P(A \cap B) = P(A)P(B).$$

Proof

The multiplicative law follows directly from Definition 2.9, the definition of conditional probability.

Notice that the multiplicative law can be extended to find the probability of the intersection of any number of events. Thus, twice applying Theorem 2.5, we obtain

$$P(A \cap B \cap C) = P[(A \cap B) \cap C] = P(A \cap B)P(C|A \cap B)$$
$$= P(A)P(B|A)P(C|A \cap B).$$

The probability of the intersection of any number of, say, k events can be obtained in the same manner:

$$P(A_1 \cap A_2 \cap A_3 \cap \cdots \cap A_k) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_k|A_1 \cap A_2 \cap \cdots \cap A_{k-1}).$$

The additive law of probability gives the probability of the union of two events.

THEOREM **2.6**

The Additive Law of Probability The probability of the union of two events A and B is

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

If A and B are mutually exclusive events, $P(A \cap B) = 0$ and

$$P(A \cup B) = P(A) + P(B).$$

Proof

The proof of the additive law can be followed by inspecting the Venn diagram in Figure 2.10.

Notice that $A \cup B = A \cup (\overline{A} \cap B)$, where A and $(\overline{A} \cap B)$ are mutually exclusive events. Further, $B = (\overline{A} \cap B) \cup (A \cap B)$, where $(A \cap \overline{B})$ and $(A \cap B)$ are mutually exclusive events. Then, by Axiom 3,

$$P(A \cup B) = P(A) + P(\overline{A} \cap B) \quad \text{and} \quad P(B) = P(\overline{A} \cap B) + P(A \cap B).$$

The equality given on the right implies that $P(\overline{A} \cap B) = P(B) - P(A \cap B)$. Substituting this expression for $P(\overline{A} \cap B)$ into the expression for $P(A \cup B)$ given in the left-hand equation of the preceding pair, we obtain the desired result:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

The probability of the union of three events can be obtained by making use of Theorem 2.6. Observe that

$$\begin{aligned} P(A \cup B \cup C) &= P[A \cup (B \cup C)] \\ &= P(A) + P(B \cup C) - P[A \cap (B \cup C)] \\ &= P(A) + P(B) + P(C) - P(B \cap C) - P[(A \cap B) \cup (A \cap C)] \\ &= P(A) + P(B) + P(C) - P(B \cap C) - P(A \cap B) - P(A \cap C) \\ &\quad + P(A \cap B \cap C) \end{aligned}$$

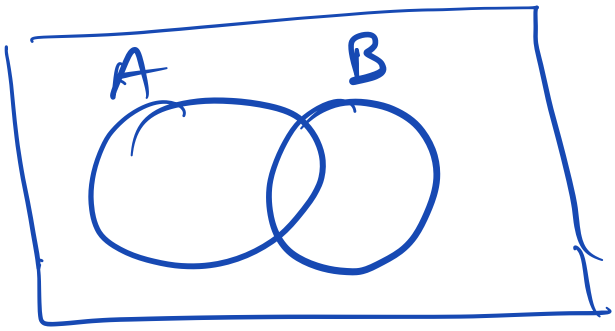
because $(A \cap B) \cap (A \cap C) = A \cap B \cap C$.

- Another useful result expressing the relationship between the probability of an event & its compliment

↳ We can use axioms.

↕

Example



If A is an event, then:

$$\mathbb{P}(A) = 1 - \mathbb{P}(\bar{A})$$

Proof:

$S = A \cup \bar{A}$. Because A & $\bar{A}$ are mutually exclusive events, it follows that $\mathbb{P}(S) = \mathbb{P}(A) + \mathbb{P}(\bar{A})$

$$\therefore \boxed{\mathbb{P}(A) + \mathbb{P}(\bar{A}) = 1}$$

Note: Sometimes its easier to calculate $\mathbb{P}(\bar{A})$ than $\mathbb{P}(A)$.

↳ Hence we can use the above relationship.

2.9 - Calculating the Probability of an Event:

Event-Composition Method

- We have learned before that sets (events) can be often be expressed as unions, intersections, or complements of other sets

- The Event-Composition Method for calculating the probability of an event, A, expresses A as a composition involving unions &/or intersections

→ The laws of probability are then applied to find P(A)

EXAMPLE 2.17 Of the voters in a city, 40% are Republicans and 60% are Democrats. Among the Republicans 70% are in favor of a bond issue, whereas 80% of the Democrats favor the issue. If a voter is selected at random in the city, what is the probability that he or she will favor the bond issue?

Solution Let F denote the event "favor the bond issue," R the event "a Republican is selected," and D the event "a Democrat is selected." Then $P(R) = .4$, $P(D) = .6$, $P(F|R) = .7$, and $P(F|D) = .8$. Notice that

$$P(F) = P[(F \cap R) \cup (F \cap D)] = P(F \cap R) + P(F \cap D)$$

because $(F \cap R)$ and $(F \cap D)$ are mutually exclusive events. Figure 2.11 will help you visualize the result that $F = (F \cap R) \cup (F \cap D)$. Now

$$P(F \cap R) = P(F|R)P(R) = (.7)(.4) = .28,$$

$$P(F \cap D) = P(F|D)P(D) = (.8)(.6) = .48.$$

It follows that

$$P(F) = .28 + .48 = .76.$$

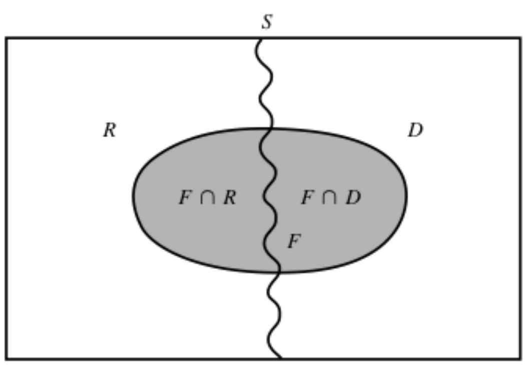


FIGURE **2.11** Venn diagram for events of Example 2.17

Note :

In General: [Event-Composition method]

A summary of the steps used in the event-composition method follows:

1. Define the experiment.
2. Visualize the nature of the sample points. Identify a few to clarify your thinking.
3. Write an equation expressing the event of interest—say, A—as a composition of two or more events, using unions, intersections, and/or complements. (Notice that this equates point sets.) Make certain that event A and the event implied by the composition represent the same set of sample points.
4. Apply the additive and multiplicative laws of probability to the compositions obtained in step 3 to find $P(A)$.

* Step 3: is the most difficult because we can form many compositions; that will be equivalent to event A.

Trick: Form a composition in which all the probabilities appearing in step 4 are known.

* This method does not require listing the sample points in S, but it does require a clear understanding of the nature of typical sample point.

* <u>Error students make</u> :

↳ Occurs in writing the composition.

↳ Always test your equality to make certain that the composition implies an event that contains the same set of sample points as those in A.

2.10 - The Law of Total Probability & Bayes' Rule

→ The event-composition approach is sometimes facilitated by viewing the sample space, S as a union of mutually exclusive subsets & using the law of probability

DEFINITION 2.11 For some positive integer k, let the sets $B_1, B_2, \ldots, B_k$ be such that

1. $S = B_1 \cup B_2 \cup \cdots \cup B_k$.
2. $B_i \cap B_j = \emptyset$, for $i \neq j$.

Then the collection of sets $\{B_1, B_2, \ldots, B_k\}$ is said to be a *partition* of S.

If A is any subset of S and $\{B_1, B_2, \ldots, B_k\}$ is a partition of S, A can be *decomposed* as follows:

$$A = (A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_k).$$

Figure 2.12 illustrates this decomposition for $k = 3$.

THEOREM 2.8 Assume that $\{B_1, B_2, \ldots, B_k\}$ is a partition of S (see Definition 2.11) such that $P(B_i) > 0$, for $i = 1, 2, \ldots, k$. Then for any event A

$$P(A) = \sum_{i=1}^{k} P(A|B_i)P(B_i).$$

Proof Any subset A of S can be written as

$$\begin{aligned} A &= A \cap S = A \cap (B_1 \cup B_2 \cup \cdots \cup B_k) \\ &= (A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_k). \end{aligned}$$

Notice that, because $\{B_1, B_2, \cdots, B_k\}$ is a partition of S, if $i \neq j$,

$$(A \cap B_i) \cap (A \cap B_j) = A \cap (B_i \cap B_j) = A \cap \emptyset = \emptyset$$

and that $(A \cap B_i)$ and $(A \cap B_j)$ are mutually exclusive events. Thus,

$$\begin{aligned} P(A) &= P(A \cap B_1) + P(A \cap B_2) + \cdots + P(A \cap B_k) \\ &= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \cdots + P(A|B_k)P(B_k) \\ &= \sum_{i=1}^{k} P(A|B_i)P(B_i). \end{aligned}$$

In the examples and exercises that follow, you will see that it is sometimes much easier to calculate the conditional probabilities $P(A|B_i)$ for suitably chosen B_i than it is to compute $P(A)$ directly. In such cases, the law of total probability can be applied

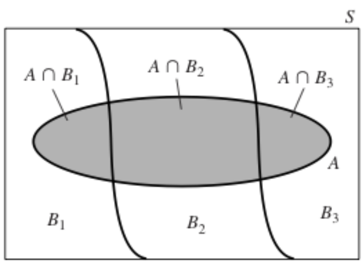


FIGURE 2.12 Decomposition of event A

* Note:
sometimes it is easier to calculate the conditional probabilities $P(A|B_i)$ for suitably chosen B_i, than it is to compute $P(A)$ directly

- Using Theorem 2.8 → derive Bayes' rule

THEOREM 2.9 **Bayes' Rule** Assume that $\{B_1, B_2, \ldots, B_k\}$ is a partition of S (see Definition 2.11) such that $P(B_i) > 0$, for $i = 1, 2, \ldots, k$. Then

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^{k} P(A|B_i)P(B_i)}.$$

Proof The proof follows directly from the definition of conditional probability and the law of total probability. Note that

$$P(B_j|A) = \frac{P(A \cap B_j)}{P(A)} = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^{k} P(A|B_i)P(B_i)}.$$

Example #

EXAMPLE **2.23** An electronic fuse is produced by five production lines in a manufacturing operation. The fuses are costly, are quite reliable, and are shipped to suppliers in 100-unit lots. Because testing is destructive, most buyers of the fuses test only a small number of fuses before deciding to accept or reject lots of incoming fuses.

All five production lines produce fuses at the same rate and normally produce only 2% defective fuses, which are dispersed randomly in the output. Unfortunately, production line 1 suffered mechanical difficulty and produced 5% defectives during the month of March. This situation became known to the manufacturer after the fuses had been shipped. A customer received a lot produced in March and tested three fuses. One failed. What is the probability that the lot was produced on line 1? What is the probability that the lot came from one of the four other lines?

Solution Let B denote the event that a fuse was drawn from line 1 and let A denote the event that a fuse was defective. Then it follows directly that

$$P(B) = 0.2 \quad \text{and} \quad P(A|B) = 3(.05)(.95)^2 = .135375.$$

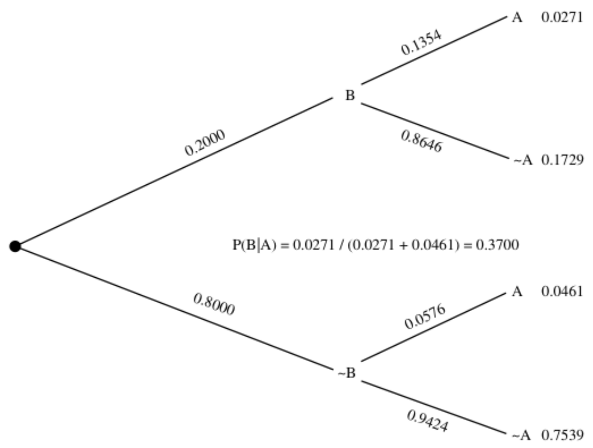


FIGURE **2.13** Tree diagram for calculations in Example 2.23. $\sim A$ and $\sim B$ are alternative notations for $\overline{A}$ and $\overline{B}$, respectively.

Similarly,

$$P(\overline{B}) = 0.8 \quad \text{and} \quad P(A|\overline{B}) = 3(.02)(.98)^2 = .057624.$$

Note that these conditional probabilities were very easy to calculate. Using the law of total probability,

$$\begin{aligned} P(A) &= P(A|B)P(B) + P(A|\overline{B})P(\overline{B}) \\ &= (.135375)(.2) + (.057624)(.8) = .0731742. \end{aligned}$$

Finally,

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{P(A|B)P(B)}{P(A)} = \frac{(.135375)(.2)}{.0731742} = .37,$$

and

$$P(\overline{B}|A) = 1 - P(B|A) = 1 - .37 = .63.$$

Figure 2.13, obtained using the applet *Bayes' Rule as a Tree*, illustrates the various steps in the computation of $P(B|A)$. ■

2.11- Numerical Events & Random Variables

- Let Y denote a variable to be measured in an experiment.

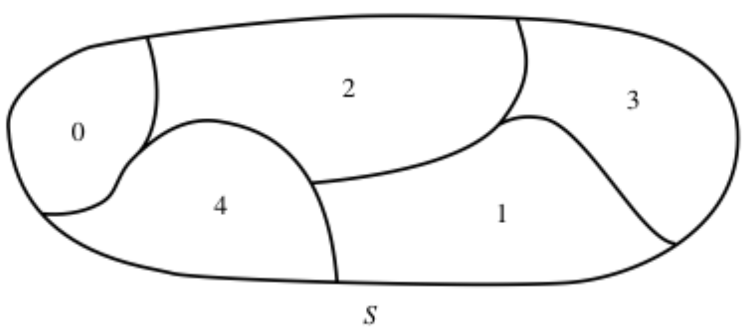
Because the value of Y will vary depending on the outcome of the experiment.

⇒ called random variable

- To each point in the sample space we will assign a real number denoting the value of the variable Y.

↳ The value assigned to Y will vary from one sample point to another, but some can have the same numerical value

FIGURE 2.14 Partitioning S into subsets that define the events $Y = 0, 1, 2, 3,$ and 4



∴ The above subsets are <u>mutually exclusive</u> since no point is assigned the same value of Y.

* A random variable is a real-valued function for which the domain is a sample space

EXAMPLE **2.24** Define an experiment as tossing two coins and observing the results. Let Y equal the number of heads obtained. Identify the sample points in S, assign a value of Y to each sample point, and identify the sample points associated with each value of the random variable Y.

Solution Let H and T represent head and tail, respectively; and let an ordered pair of symbols identify the outcome for the first and second coins. (Thus, HT implies a head on the first coin and a tail on the second.) Then the four sample points in S are E_1: HH, E_2: HT, E_3: TH and E_4: TT. The values of Y assigned to the sample points depend on the number of heads associated with each point. For E_1: HH, two heads were observed, and E_1 is assigned the value $Y = 2$. Similarly, we assign the values $Y = 1$ to E_2 and E_3 and $Y = 0$ to E_4. Summarizing, the random variable Y can take three values, $Y = 0$, 1, and 2, which are events defined by specific collections of sample points:

$$\{Y = 0\} = \{E_4\}, \qquad \{Y = 1\} = \{E_2, E_3\}, \qquad \{Y = 2\} = \{E_1\}. \qquad \blacksquare$$

Let y denote an observed value of the random variable Y. Then $P(Y = y)$ is the sum of the probabilities of the sample points that are assigned the value y.

EXAMPLE **2.25** Compute the probabilities for each value of Y in Example 2.24.

Solution The event $\{Y = 0\}$ results only from sample point E_4. If the coins are balanced, the sample points are equally likely; hence,

$$P(Y = 0) = P(E_4) = 1/4.$$

Similarly,

$$P(Y = 1) = P(E_2) + P(E_3) = 1/2 \quad \text{and} \quad P(Y = 2) = P(E_1) = 1/4. \qquad \blacksquare$$

A more detailed examination of random variables will be undertaken in the next two chapters.

2.13 - Random Sampling

- A statistical experiment involves the observation of a sample, selected from a larger body of data, existing or conceptual called population
  - ↳ The measurements in the sample, viewed as observations of the values of one or more random variables, are then employed to make an inference about the characteristics of the target population

* In general 2 sample selection
  - sampling with replacement
  - sampling without replacement

Note: "design of an experiment" is method of sampling.

↳ Affects both quantity of information in the sample & probability of observation

- Random sample

DEFINITION 2.13

Let N and n represent the numbers of elements in the population and sample, respectively. If the sampling is conducted in such a way that each of the $\binom{N}{n}$ samples has an equal probability of being selected, the sampling is said to be random, and the result is said to be a *random sample*.

2.13 Summary

This chapter has been concerned with providing a model for the repetition of an experiment and, consequently, a model for the population frequency distributions of Chapter 1. The acquisition of a probability distribution is the first step in forming a theory to model reality and to develop the machinery for making inferences.

An experiment was defined as the process of making an observation. The concepts of an event, a simple event, the sample space, and the probability axioms have provided a probabilistic model for calculating the probability of an event. Numerical events and the definition of a random variable were introduced in Section 2.11.

Inherent in the model is the sample-point approach for calculating the probability of an event (Section 2.5). Counting rules useful in applying the sample-point method were discussed in Section 2.6. The concept of conditional probability, the operations of set algebra, and the laws of probability set the stage for the event-composition method for calculating the probability of an event (Section 2.9).

Of what value is the theory of probability? It provides the theory and the tools for calculating the probabilities of numerical events and hence the probability

distributions for the random variables that will be discussed in Chapter 3. The numerical events of interest to us appear in a sample, and we will wish to calculate the probability of an observed sample to make an inference about the target population. Probability provides both the foundation and the tools for statistical inference, the objective of statistics.

References and Further Readings

Cramer, H. 1973. *The Elements of Probability Theory and Some of Its Applications*, 2d ed. Huntington, N.Y.: Krieger.

Feller, W. 1968. *An Introduction to Probability Theory and Its Applications*, 3d ed., vol. 1. New York: Wiley.

———. 1971. *An Introduction to Probability Theory and Its Applications*, 2d ed., vol. 2. New York: Wiley.

Meyer, P. L. 1970. *Introductory Probability and Statistical Applications*, 2d ed. Reading, Mass.: Addison-Wesley.

Parzen, E. 1992. *Modern Probability Theory and Its Applications*. New York: Wiley-Interscience.

Riordan, J. 2002. *Introduction to Combinatorial Analysis*. Mineola, N.Y.: Dover Publications.